

BAB III

METODE *CHI-SQUARED AUTOMATIC INTERACTION DETECTION*

3.1 Analisis CHAID

Metode CHAID pertama kali diperkenalkan G. V. Kass 1980, metode CHAID merupakan teknik yang lebih awal dikenal sebagai *Automatic Interaction Detection (AID)*. Metode CHAID secara umum bekerja dengan mempelajari hubungan antara variabel dependen dengan beberapa variabel independen, kemudian mengklasifikasi sampel berdasarkan hubungan tersebut.

CHAID adalah sebuah metode untuk mengklasifikasikan data kategori dimana tujuan prosedurnya adalah untuk membagi rangkaian data menjadi subgrup-subgrup berdasarkan pada variabel dependennya (Lehmann dan Eherler, 2001). Hasil dari pengklasifikasian dalam CHAID akan ditampilkan dalam sebuah diagram pohon. Menurut Gallagher (2000), CHAID merupakan suatu teknik iteratif yang menguji satu-persatu variabel independen yang digunakan dalam klasifikasi, dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya.

Menurut Kunto dan Hasana (2006), CHAID digunakan untuk membentuk segmentasi yang membagi sebuah sampel menjadi dua atau lebih kelompok yang berbeda berdasarkan sebuah kriteria tertentu. Hal ini kemudian diteruskan dengan membagi kelompok-kelompok tersebut menjadi kelompok yang lebih kecil berdasarkan variabel-variabel independen yang lain. Proses tersebut terus berlanjut sampai tidak ditemukan lagi variabel-variabel independen yang signifikan secara statistik.

Dari uraian diatas dapat disimpulkan CHAID adalah metode untuk mengklasifikasikan data yang membagi sampel menjadi dua atau lebih kelompok berdasarkan kriteria tertentu dan hasil pengklasifikasiannya ditampilkan dalam diagram pohon.

3.2 Variabel-variabel dalam Analisis CHAID

Variabel dependen dan independen dalam analisis CHAID merupakan variabel kategori. Menurut Gallegher (2000), CHAID akan membedakan variabel-variabel independen kategori menjadi tiga bentuk yang berbeda yaitu **monotonik, bebas, mengambang**.

1. Monotonik (*Monotonic*)

Kategori-kategori pada variabel ini dapat dikombinasikan atau digabungkan hanya jika keduanya berdekatan satu sama lain, yaitu variabel-variabel yang kategorinya mengikuti urutan aslinya (data ordinal). Contohnya adalah: usia dan pendapatan.

2. Bebas (*Free*)

Kategori-kategori pada variabel ini dapat dikombinasikan atau digabungkan ketika keduanya berdekatan ataupun tidak (data nominal). Contohnya: pekerjaan, kelompok etnik, dan area geografis.

3. Mengambang (*Floating*)

Kategori-kategori pada variabel ini dapat diperlakukan seperti monotonik kecuali untuk kategori yang *missing value*, yang dapat berkombinasi dengan kategori manapun.

3.3 Algoritma CHAID

Algoritma CHAID digunakan untuk melakukan pemisahan dan penggabungan kategori-kategori dalam variabel yang dipakai dalam analisisnya. Secara garis besar algoritma ini dapat dibagi menjadi tiga tahap, yaitu Penggabungan (*merging*), Pemisahan (*splitting*) dan Penghentian (*stopping*). Diagram pohon dimulai dari *root node* (node akar) melalui tiga tahap tersebut pada setiap simpul (*node*) yang terbentuk dan secara berulang.

3.3.1 Tahap penggabungan

Ayu Wulandary, 2014

Klasifikasi keputusan nasabah untuk menggunakan ATM dengan metode Chi-Square Automatic Interaction Detection (CHAID)

Universitas Pendidikan Indonesia | repository.upi.edu

Tahap pertama algoritma CHAID adalah penggabungan. Pada tahap ini akan diperiksa signifikansi dari masing-masing kategori variabel independen terhadap variabel dependen. Tahap penggabungan untuk setiap independen dalam menggabungkan kategori-kategori nonsignifikan adalah sebagai berikut:

1. Bentuk tabel kontingensi dua arah untuk masing-masing variabel independen dengan variabel dependennya.
2. Hitung statistik *chi-square* untuk setiap pasang kategori yang dapat dipilih untuk digabungkan menjadi satu, untuk menguji kebebasannya dalam sebuah sub tabel kontingensi $2 \times j$ yang dibentuk oleh sepasang kategori tersebut dengan variabel dependennya yang mempunyai sebanyak j kategori. Langkah uji *chi square* adalah sebagai berikut:

- a. Perumusan hipotesisnya

H_0 : tidak terdapat hubungan antara kategori i pada variabel independen dan kategori j pada variabel dependen

H_1 : terdapat hubungan kategori i pada variabel independen dan kategori j pada variabel dependen

- b. Besaran-besaran yang diperlukan

Menghitung E_{ij}

- c. Statistik uji

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- d. Kriteria Pengujian

H_0 di tolak, jika $\chi^2 \geq \chi^2_{\alpha; (b-1)(k-1)}$

- e. Kesimpulan

Penaksiran dari H_0 diterima atau ditolak.

3. Untuk masing-masing nilai *chi-square* berpasangan, hitung *p-value* berpasangan bersamaan. Diantara pasangan-pasangan yang tidak signifikan, gabungkan sebuah pasangan kategori yang paling mirip (yaitu

pasangan yang mempunyai nilai *chi-square* berpasangan terkecil dan *p-value* terbesar) menjadi sebuah kategori tunggal dan kemudian lanjutkan ke langkah nomor 4.

4. Periksa kembali kesignifikansian kategori baru setelah digabungkan dengan kategori lainnya dalam variabel independen. Jika masih ada pasangan yang belum signifikan, ulangi langkah nomor 3. Jika semua signifikan lanjutkan ke langkah selanjutnya.
5. Hitung *p-value* terkoreksi Bonferroni didasarkan pada tabel yang telah digabung.

3.3.2 Tahap Pemisahan (*Splitting*)

Langkah pemisahan digunakan untuk memilih variabel independen mana yang akan digunakan untuk pembagian titik terbaik. Pemisahan dilakukan dengan membandingkan *p-value* (dari tahap penggabungan) pada setiap variabel independen. Langkah pembagian adalah sebagai berikut:

1. Pilih variabel independen yang memiliki *p-value* terkecil (paling signifikan).
2. Jika *p-value* kurang dari sama dengan tingkat spesifikasi alpha, pemisahan simpul menggunakan variabel independen ini. Jika tidak ada variabel independen dengan nilai *p-value* yang signifikan, tidak dilakukan pemisahan dan *simpul* ditentukan sebagai terminal simpul (simpul akhir).

3.3.3 Tahap Penghentian (*Stopping*)

Tahap penghentian dilakukan jika suatu proses pertumbuhan pohon harus dihentikan sesuai dengan peraturan penghentian dibawah ini:

1. Tidak ada lagi variabel independen yang signifikan menunjukkan perbedaan terhadap variabel dependen.
2. Jika pohon sekarang mencapai batas nilai maksimum pohon dari spesifikasi, maka proses pertumbuhan akan berhenti. Misalkan ditetapkan

batas kedalaman pertumbuhan pohon klasifikasi adalah 3, ketika pertumbuhan pohon sudah mencapai kedalaman 3 maka pertumbuhan pohon klasifikasi dihentikan.

3. Jika ukuran dari *child node* kurang dari nilai ukuran *child node* minimum spesifikasi, atau berisi pengamatan-pengamatan dengan banyak yang terlalu sedikit maka *node* tidak akan di-*split*. Misalkan ditetapkan ukuran minimal *child node* adalah 10, ketika *splitting* menghasilkan ukuran *child node* kurang dari 10, maka *node* tersebut tidak akan dipecah.

3.4 Koreksi Bonferroni

Andaikan bahwa variabel independen memiliki c kategori, setelah dilakukan penggabungan dikurangi menjadi r kategori, maka perkalian Bonferroni adalah banyaknya cara yang mungkin yang mana c kategori dapat digabungkan menjadi r kategori. Dengan demikian nilai p -value dari uji *chi-square* untuk independensi yang baru merupakan perkaliannya dengan pengali Bonferroni sesuai dengan jenis variabelnya (Gallagher, 2000).

Koreksi Bonferroni adalah suatu proses koreksi yang digunakan ketika beberapa uji statistik untuk kebebasan atau ketidakbebasan dilakukan secara bersamaan (Kunto dan Hasana, 2006). Koreksi Bonferroni biasanya digunakan dalam perbandingan berganda.

Gallagher (2000) menyebutkan bahwa pengali *Bonferroni* untuk masing-masing jenis variabel independen adalah sebagai berikut:

1. Variabel independen monotonik

$$M = \binom{c-1}{r-1} \quad \dots (3.1)$$

dimana :

M = pengali Bonferroni

c = banyaknya kategori variabel independen awal

r = banyaknya kategori variabel independen setelah penggabungan

2. Variabel independen bebas

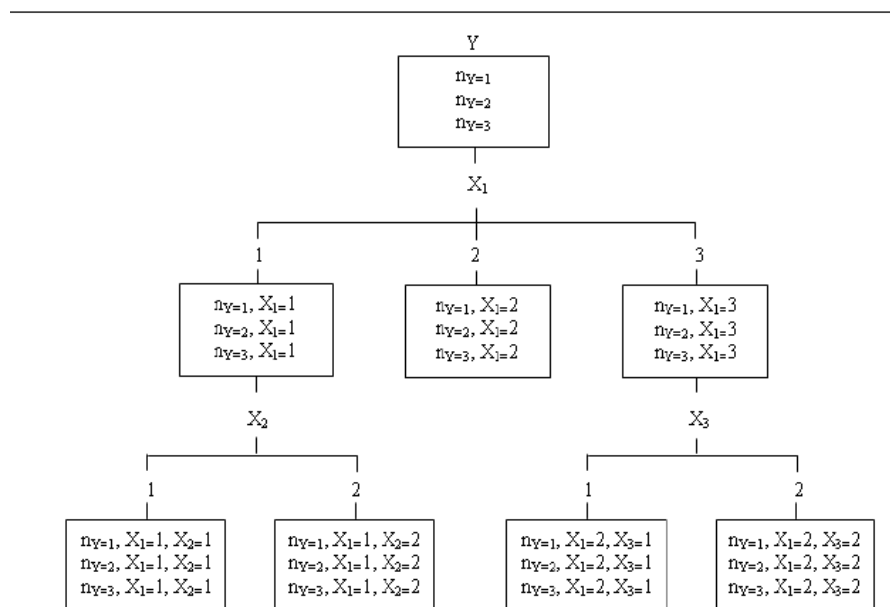
$$M = \sum_{i=0}^{r-1} (-1)^i \frac{(r-1)^c}{i! (r-i)!} \quad \dots (3.2)$$

3. Variabel independen mengambang

$$M = \binom{c-2}{r-2} + r \binom{c-2}{r-2} \quad \dots (3.3)$$

3.5 Diagram Pohon Klasifikasi CHAID

CHAID akan menghasilkan sebuah diagram pohon klasifikasi yang menggambarkan pembentukan segmen. Diagram pohon CHAID ditunjukkan pada Gambar 3.1 Diagram CHAID terdiri dari batang pohon (*tree trunk*) dengan membagi (*split*) menjadi lebih kecil berupa cabang-cabang (*branches*).



Gambar 3.1
Diagram Pohon CHAID

Menurut Myers (Kunto dan Hasana, 2006), diagram pohon CHAID mengikuti aturan “dari atas ke bawah” (*Top-down stopping rule*), dimana diagram pohon disusun mulai dari kelompok induk (*parent node*), berlanjut di bawahnya sub kelompok (*child node*) yang berturut-turut dari hasil pembagian kelompok induk berdasarkan kriteria tertentu. Node pada ujung pohon yang tidak terdapat percabangan lagi disebut *terminal node*. Tiap-tiap *node* dari diagram pohon ini menggambarkan sub kelompok dari sampel yang diteliti dan berisi keseluruhan sampel dan frekuensi absolut ni untuk setiap kategori yang disusun.

Pada pohon klasifikasi CHAID terdapat istilah kedalaman (*depth*) yang berarti banyak tingkatan *node-node* sub kelompok sampai ke bawah pada *node* sub kelompok yang terakhir. Pada kedalaman pertama, sampel dibagi oleh X_1 sebagai variabel independen terbaik untuk variabel dependen berdasarkan uji *chi-square*. Tiap *node* berisi informasi tentang frekuensi variabel Y, sebagai variabel dependen, yang merupakan bagian dari sub kelompok yang dihasilkan berdasarkan kategori yang disebutkan (X_1). Pada kedalaman $ke - 2$ (*node* X_2 dan X_3) merupakan pembagian dari X_1 (untuk *node* $ke - 1$ dan $ke - 3$). Dengan cara yang sama, sampel selanjutnya dibagi oleh variabel independen yang lain, yaitu X_2 dan X_3 , dan selanjutnya menjadi sub kelompok pada *node* ke-4, 5, 6, dan 7 (Lehmann dan Eherler, 2001). Pada masing-masing *node* ditampilkan persentase responden untuk setiap kategori dari variabel dependen, dan juga ditunjukkan jumlah total responden untuk masing-masing *node*.

3.6 Contoh Kasus Pembentukan Pohon Keputusan dengan Algoritma CHAID

Data keputusan bermain

No	Outlook	Temperatur	Plays
1	sunny	Hot	yes

Ayu Wulandary, 2014

Klasifikasi keputusan nasabah untuk menggunakan ATM dengan metode Chi-Square Automatic Interaction Detection (CHAID)

Universitas Pendidikan Indonesia | repository.upi.edu

2	<i>overcast</i>	<i>Cold</i>	<i>yes</i>
3	<i>overcast</i>	<i>Cold</i>	<i>yes</i>
4	<i>sunny</i>	<i>Mild</i>	<i>yes</i>
5	<i>overcast</i>	<i>Cold</i>	<i>yes</i>
6	<i>overcast</i>	<i>Mild</i>	<i>yes</i>
7	<i>overcast</i>	<i>Cold</i>	<i>yes</i>
8	<i>overcast</i>	<i>Mild</i>	<i>yes</i>
9	<i>rainy</i>	<i>Cold</i>	<i>no</i>
10	<i>sunny</i>	<i>Cold</i>	<i>no</i>
11	<i>rainy</i>	<i>Hot</i>	<i>no</i>
12	<i>sunny</i>	<i>Cold</i>	<i>yes</i>

Lanjutan

No	<i>Outlook</i>	<i>Temperatur</i>	<i>Plays</i>
13	<i>sunny</i>	<i>Hot</i>	<i>no</i>
14	<i>sunny</i>	<i>Hot</i>	<i>yes</i>

- ❖ Variabel dependen Keputusan bermain. Skala data untuk variabel ini adalah nominal, dibedakan menjadi dua kategori yaitu *yes* dengan kode 1 dan *no* dengan kode 2.
- ❖ Variabel Independen *outlook*, skala data untuk variabel ini adalah nominal, dibedakan menjadi tiga kategori yaitu *sunny* dengan kode 1, *overcast* dengan kode 2 dan *rainy* dengan kode 3. Variabel Independen Temperatur, skala data untuk variabel ini adalah nominal, dibedakan menjadi tiga kategori yaitu *hot* dengan kode 1, *mild* dengan kode 2 dan *cold* dengan kode 3.

Bentuk tabel kontingensi dua arah untuk masing-masing variabel.

Perumusan hipotesisnya sebagai berikut:

Ayu Wulandary, 2014

Klasifikasi keputusan nasabah untuk menggunakan ATM dengan metode Chi-Square Automatic Interaction Detection (CHAID)

Universitas Pendidikan Indonesia | repository.upi.edu

- ❖ H_0 : Tidak terdapat hubungan antara variabel independen dengan variabel dependen
- ❖ H_1 : Terdapat hubungan antara variabel independen dengan variabel dependen.

Variabel independen *outlook**Y

Tabel 3.1
Tabel Frekuensi Teramati 1

		O		Total
		Y = yes	Y = no	
<i>Outlook</i>	Sunny	4	2	6
	Overcast	6	0	6
	Total	10	2	12

Frekuensi sel yang diharapkan dapat dicari dengan cara sebagai berikut

$$E_{11} = \left(\frac{n_{1.} \cdot n_{.1}}{n} \right) = \frac{60}{12} = 5$$

$$E_{12} = \left(\frac{n_{1.} \cdot n_{.2}}{n} \right) = \frac{12}{12} = 1$$

$$E_{21} = \left(\frac{n_{2.} \cdot n_{.1}}{n} \right) = \frac{60}{12} = 5$$

$$E_{22} = \left(\frac{n_{2.} \cdot n_{.2}}{n} \right) = \frac{12}{12} = 1$$

Nilai statistik uji χ^2 diperoleh sebagai berikut

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{(4 - 5)^2}{5} + \frac{(2 - 1)^2}{1} + \frac{(6 - 5)^2}{5} + \frac{(0 - 1)^2}{1}$$

$$\chi^2 = 2,4$$

Dengan bantuan Excel diperoleh nilai $p = 0,12$

Tabel 3.2
Tabel Frekuensi Teramati 2

	O	Total
--	---	-------

		Y = yes	Y = no	
<i>Outlook</i>	Sunny	4	2	6
	Rainy	0	2	2
	Total	4	4	8

Frekuensi sel yang diharapkan dapat dicari dengan cara sebagai berikut

$$E_{11} = \left(\frac{n_{1.} \cdot n_{.1}}{n} \right) = \frac{24}{8} = 3$$

$$E_{12} = \left(\frac{n_{1.} \cdot n_{.2}}{n} \right) = \frac{24}{8} = 2$$

$$E_{21} = \left(\frac{n_{2.} \cdot n_{.1}}{n} \right) = \frac{8}{8} = 1$$

$$E_{22} = \left(\frac{n_{2.} \cdot n_{.2}}{n} \right) = \frac{8}{8} = 1$$

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{(4 - 3)^2}{3} + \frac{(2 - 3)^2}{3} + \frac{(0 - 1)^2}{1} + \frac{(2 - 1)^2}{1}$$

$$\chi^2 = 2,67$$

Dengan bantuan Excel diperoleh nilai $p = 0,10$

Tabel 3.3
Tabel Frekuensi Teramati 3

		O		Total
		Y = yes	Y = no	
<i>Outlook</i>	overcast	6	0	6
	Rainy	0	2	2
	Total	6	2	8

Frekuensi sel yang diharapkan dapat dicari dengan cara sebagai berikut

$$E_{11} = \left(\frac{n_{1.} \cdot n_{.1}}{n} \right) = \frac{36}{8} = 4,5$$

$$E_{12} = \left(\frac{n_{1.} \cdot n_{.2}}{n} \right) = \frac{12}{8} = 1,5$$

$$E_{21} = \left(\frac{n_{2.}n_{.1}}{n} \right) = \frac{12}{8} = 1,5$$

$$E_{22} = \left(\frac{n_{2.}n_{.2}}{n} \right) = \frac{4}{8} = 0,5$$

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{(6 - 4,5)^2}{4,5} + \frac{(0 - 1,5)^2}{1,5} + \frac{(0 - 1,5)^2}{1,5} + \frac{(2 - 0,5)^2}{0,5}$$

$$\chi^2 = 8$$

Dengan bantuan Excel diperoleh nilai $p = 0,0047$

- ❖ Gabungkan sebuah pasangan kategori yang paling mirip yaitu pasangan yang mempunyai nilai χ^2 terkecil menjadi satu kategori. Kategori *sunny* dan *overcast* yang digabungkan menjadi satu.

Tabel 3.4
Tabel Frekuensi Teramati 4

		O		Total
		Y = yes	Y = no	
<i>Outlook</i>	Sunny,Overcast	10	2	12
	Rainy	0	2	2
	Total	10	4	14

Periksa kesignifikansian katategori baru setelah digabungkan

$$E_{11} = \left(\frac{n_{1.}n_{.1}}{n} \right) = \frac{120}{14} = 8,6$$

$$E_{12} = \left(\frac{n_{1.}n_{.2}}{n} \right) = \frac{48}{14} = 3,4$$

$$E_{21} = \left(\frac{n_{2.}n_{.1}}{n} \right) = \frac{20}{14} = 1,4$$

$$E_{22} = \left(\frac{n_{2.}n_{.2}}{n} \right) = \frac{8}{14} = 0,6$$

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{(12 - 8,6)^2}{8,6} + \frac{(2 - 3,4)^2}{3,4} + \frac{(0 - 1,4)^2}{1,4} + \frac{(2 - 0,6)^2}{0,6}$$

$$\chi^2 = 5,83$$

Dari Tabel Distribusi Chi-Kuadrat dengan peluang 0,95 dan dk = 1 diperoleh $\chi^2_{0,95,1} = 3,84$. Karena $5,83 > 3,84$ maka H_0 ditolak artinya terdapat hubungan antara variabel independen *outlook* dan keputusan bermain.

Variabel Independen Temperatur*Y

Tabel 3.5
Tabel Frekuensi Teramati 5

		O		Total
		Y = yes	Y = no	
Temperatur	hot	2	2	4
	mild	3	0	3
	Total	5	2	7

Frekuensi sel yang diharapkan dapat dicari dengan cara sebagai berikut

$$E_{11} = \left(\frac{n_{1.} \cdot n_{.1}}{n} \right) = \frac{20}{7} = 2,86$$

$$E_{12} = \left(\frac{n_{1.} \cdot n_{.2}}{n} \right) = \frac{8}{7} = 1,14$$

$$E_{21} = \left(\frac{n_{2.} \cdot n_{.1}}{n} \right) = \frac{15}{7} = 2,14$$

$$E_{22} = \left(\frac{n_{2.} \cdot n_{.2}}{n} \right) = \frac{6}{7} = 0,86$$

Nilai statistik uji χ^2 diperoleh sebagai berikut

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Ayu Wulandary, 2014

Klasifikasi keputusan nasabah untuk menggunakan ATM dengan metode Chi-Square Automatic Interaction Detection (CHAID)

Universitas Pendidikan Indonesia | repository.upi.edu

$$\chi^2 = \frac{(2 - 2,86)^2}{2,86} + \frac{(2 - 1,14)^2}{1,14} + \frac{(3 - 2,14)^2}{2,14} + \frac{(0 - 0,86)^2}{0,86}$$

$$\chi^2 = 2,1$$

Tabel 3.6
Tabel Frekuensi Teramati 6

		O		Total
		Y = yes	Y = no	
Temperatur	hot	2	2	4
	cold	5	2	7
	Total	7	4	11

Frekuensi sel yang diharapkan dapat dicari dengan cara sebagai berikut

$$E_{11} = \left(\frac{n_{1.} \cdot n_{.1}}{n} \right) = \frac{28}{11} = 2,54$$

$$E_{12} = \left(\frac{n_{1.} \cdot n_{.2}}{n} \right) = \frac{16}{11} = 1,45$$

$$E_{21} = \left(\frac{n_{2.} \cdot n_{.1}}{n} \right) = \frac{49}{11} = 4,45$$

$$E_{22} = \left(\frac{n_{2.} \cdot n_{.2}}{n} \right) = \frac{28}{11} = 2,54$$

Nilai statistik uji χ^2 diperoleh sebagai berikut

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{(2 - 2,54)^2}{2,54} + \frac{(2 - 1,45)^2}{1,45} + \frac{(5 - 4,45)^2}{4,45} + \frac{(2 - 2,54)^2}{2,54}$$

$$\chi^2 = 0,50$$

Tabel 3.7
Tabel Frekuensi Teramati 7

	O	Total
--	---	-------

		Y = yes	Y = no	
<i>Temperatur</i>	mild	3	0	3
	Cold	5	2	7
	Total	8	2	10

Frekuensi sel yang diharapkan dapat dicari dengan cara sebagai berikut

$$E_{11} = \left(\frac{n_{1.} \cdot n_{.1}}{n} \right) = \frac{24}{10} = 2,4$$

$$E_{12} = \left(\frac{n_{1.} \cdot n_{.2}}{n} \right) = \frac{6}{10} = 0,6$$

$$E_{21} = \left(\frac{n_{2.} \cdot n_{.1}}{n} \right) = \frac{56}{10} = 5,6$$

$$E_{22} = \left(\frac{n_{2.} \cdot n_{.2}}{n} \right) = \frac{14}{10} = 1,4$$

Nilai statistik uji χ^2 diperoleh sebagai berikut

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{(3 - 2,4)^2}{2,4} + \frac{(0 - 0,6)^2}{0,6} + \frac{(5 - 5,6)^2}{5,6} + \frac{(2 - 1,4)^2}{1,4}$$

$$\chi^2 = 0,107$$

- ❖ Gabungkan sebuah pasangan kategori yang paling mirip yaitu pasangan yang mempunyai nilai χ^2 terkecil dan *p-value* yang menjadi satu kategori. Kategori *mild* dan *cold* yang digabungkan menjadi satu.

Tabel 3.8
Tabel Frekuensi Teramati 8

		O		Total
		Y = yes	Y = no	
<i>Temperatur</i>	Mild,cold	8	2	10
	Hot	2	2	4
	Total	10	4	14

Periksa kesignifikansian kategory baru setelah digabungkan

$$E_{11} = \left(\frac{n_{1.} \cdot n_{.1}}{n} \right) = \frac{100}{14} = 7,14$$

$$E_{12} = \left(\frac{n_{1.} \cdot n_{.2}}{n} \right) = \frac{40}{14} = 2,86$$

$$E_{21} = \left(\frac{n_{2.} \cdot n_{.1}}{n} \right) = \frac{40}{14} = 2,86$$

$$E_{22} = \left(\frac{n_{2.} \cdot n_{.2}}{n} \right) = \frac{16}{14} = 1,14$$

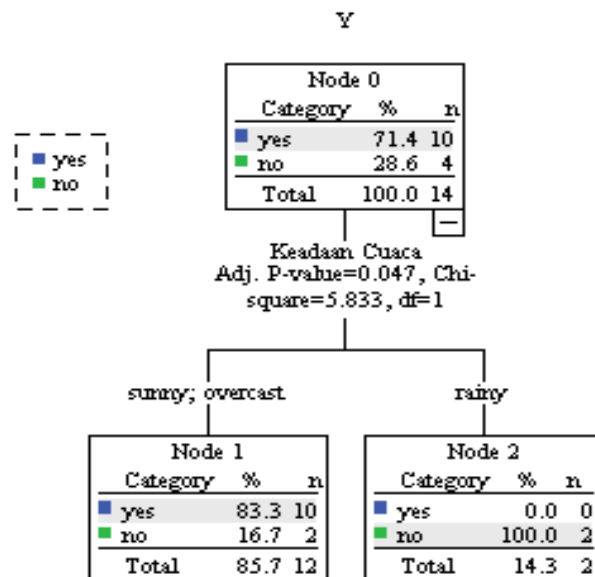
$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{(8 - 7,14)^2}{7,14} + \frac{(2 - 2,86)^2}{2,86} + \frac{(2 - 2,86)^2}{2,86} + \frac{(2 - 1,14)^2}{1,14}$$

$$\chi^2 = 1,26$$

Dari Tabel Distribusi Chi-Kuadrat dengan peluang 0,95 dan dk = 1 diperoleh $\chi^2_{0,95;1} = 3,84$. Karena $1,26 < 3,84$ maka H_0 diterima artinya tidak terdapat hubungan antara variabel independen *temperatur* dan keputusan bermain.

Berdasarkan Gambar 3.2, terlihat bahwa variabel yang berpengaruh secara signifikan dalam klasifikasi keputusan bermain adalah variabel keadaan cuaca (*outlook*). Yang menghasilkan dua simpul terminal (terminal node). Pada *outlook* sunny, overcast keputusan yang bermain sebanyak 10 orang dan keputusan tidak bermain 2 orang. Pada *outlook* rainy keputusan yang tidak bermain sebanyak 2 orang.



Gambar 3.2
Pohon Klasifikasi Contoh